



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
International Journal of Population Data Science

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa51256>

Paper:

Atkinson, M., Kennedy, J. & Brophy, S. (2017). Comparing a primary care dataset with a questionnaire based health survey. *International Journal of Population Data Science*, 1(1)
<http://dx.doi.org/10.23889/ijpds.v1i1.67>

Released under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC-BY-NC-ND).

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

International Journal of Population Data Science

Journal Website: www.ijpds.org



Comparing a primary care dataset with a questionnaire based health survey

Atkinson, Mark^{1*}, Kennedy, Jonathan¹, and Brophy, Sinead¹

¹Swansea University

Objectives

The Welsh Health Survey (WHS) has been carried out annually since 2003. Approximately 15000 adults and 3000 children are interviewed each year on a wide range of health related questions. From 2013, adult participants were asked to consent to their data being linked and incorporated into the SAIL databank, allowing linkage to the SAIL datasets. Here we focus on linkage to primary care, General Practitioner (GP), data.

This provides a unique opportunity to compare some clinical concepts which are represented in both datasets. The questionnaires in the WHS are taken at a set moment in time and are standardised for everyone questioned, and are more complete. In comparison the GP data collects responses over the patients' interactions with the GP, but the information may have been elicited in different ways and data are much more likely to be missing.

Approach

We have chosen a small number of variables scored in both datasets for comparison. BMI and its core variables height and weight are the principal numeric variables found in both datasets. Smoking behaviour is coded for 96% of people in the GP data, alcohol consumption for 79% and exercise for 62%. In contrast, in the WHS data, coding is for 99% (smoking), 99% (alcohol) and 97% (exercise).

Results

The 2013 WHS had 4362 participants linked to the SAIL databank, and 7332 from the 2014 WHS. Of these, 95% had either an exact match or a probabilistic match of 0.9 or above, with 6869 people (63%) of the combined WHS dataset having linked GP data.

Of those with GP data, 5997 (87%) have weights in the GP data at any date. But only 2429 (35%) have weights within 1

year before and 1 year after the questionnaire date. Of the 2429, 135 have no weight in the WHS data.

Conclusions

For those using WHS data, the data can be enhanced with a temporal dimension, while the GP data user can augment missing values from the WHS. Both the GP and WHS can gain confirmation from each other. This allows the non-randomness of missingness in the GP dataset to be assessed.

*Corresponding Author:

Email Address: M.Atkinson@Swansea.ac.uk (M. Atkinson)

